

# Nonparametric testing of conditional independence by means of the partial copula

Wicher P. Bergsma  
 London School of Economics and Political Science

January 25, 2011

## Abstract

We propose a new method to test conditional independence of two real random variables  $Y$  and  $Z$  conditionally on an arbitrary third random variable  $X$ . The partial copula is introduced, defined as the joint distribution of  $U = F_{Y|X}(Y|X)$  and  $V = F_{Z|X}(Z|X)$ . We call this transformation of  $(Y, Z)$  into  $(U, V)$  the partial copula transform. It is easy to show that if  $Y$  and  $Z$  are continuous for any given value of  $X$ , then  $Y \perp\!\!\!\perp Z|X$  implies  $U \perp\!\!\!\perp V$ . Conditional independence can then be tested by (i) applying the partial copula transform to the data points and (ii) applying a test of ordinary independence to the transformed data. In practice,  $F_{Y|X}$  and  $F_{Z|X}$  will need to be estimated, which can be done by, e.g., standard kernel methods. We show that under easily satisfied conditions, and for a very large class of test statistics for independence which includes the covariance, Kendall's tau, and Hoeffding's test statistic, the effect of this estimation vanishes asymptotically. Thus, for large samples, the estimation can be ignored and we have a simple method which can be used to apply a wide range of tests of independence, including ones with consistency for arbitrary alternatives, to test for conditional independence. A simulation study indicates good small sample performance. Advantages of the partial copula approach compared to competitors seem to be simplicity and generality.

Keywords: partial copula, partial correlation

## 1 Introduction

Random variables  $Y$  and  $Z$  are conditionally independent given  $X$  if, knowing the value of  $X$ , information about  $Y$  does not provide information about  $Z$ . Following Dawid (1979), this hypothesis is denoted as  $Y \perp\!\!\!\perp Z|X$ . Conditional independence models are the building blocks of graphical models, which can be used for causal modeling (Wermuth & Cox, 1996; Lauritzen, 1996; Pearl, 2000). The literature on the topic is usually restricted to the normal and categorical cases, but recently nonparametric testing has also received a fair amount of attention (Su & White, 2007, 2008; Song, 2009; Huang, 2010; Bouezmarni, Rombouts,

& Taamouti, 2010). These papers also contain useful further references and applications. At the end of this section a brief literature review is presented as well. This paper presents the partial copula approach for the nonparametric testing of the conditional independence hypothesis for real random variables  $Y$  and  $Z$  controlling for an arbitrary random variable  $X$  based on  $n$  independent and identically distributed (iid) replications of  $(X, Y, Z)$ .

In general, compared to the testing of ordinary independence, the present problem seems a difficult one (Bergsma, 2004). In particular, if, without further assumptions, for a given value of  $X$  only one  $(Y, Z)$  observation is available, this does not tell us anything at all about the conditional association for that particular value of  $X$ . Alternatively, it can be said that the difficulty lies in the possible presence of marginal dependencies of  $Y$  on  $X$  and of  $Z$  on  $X$ : if both  $Y \perp\!\!\!\perp X$  and  $Z \perp\!\!\!\perp X$ , then  $Y \perp\!\!\!\perp Z|X$  implies  $Y \perp\!\!\!\perp Z$ . This is easily shown as follows: assuming the conditions hold,

$$\begin{aligned} P(X \in A, Y \in B) &= P(X \in A, Y \in B|X = x) \\ &= P(X \in A|X = x)P(Y \in B|X = x) = P(X \in A)P(Y \in B) \end{aligned} \quad (1)$$

for measurable subsets  $A$  and  $B$  of the sample spaces of  $Y$  and  $Z$ , an  $x$  in the sample space of  $X$ . Thus, in the absence of the marginal associations, conditional independence for  $(X, Y, Z)$  can be tested simply by performing a test of ordinary independence for  $(Y, Z)$ . In practice, of course, marginal dependencies are typically present, and conditional independence might be tested by removing (some of) the marginal dependency. Below, we describe two methods which use this idea.

Firstly, the idea of removing some of the marginal association is used to construct the well-known partial correlation coefficient. The assumption is made that  $Y = g(X) + U'$  and  $Z = h(X) + V'$  where the error terms  $U'$  and  $V'$  have zero expectation conditionally on  $X$ . Now clearly the correlations between  $X$  and  $U'$  and between  $X$  and  $V'$  are zero, i.e., replacing  $Y$  by  $U'$  and  $Z$  by  $V'$  removes some of the marginal association. The partial correlation coefficient is the correlation between  $U'$  and  $V'$ , which is zero under conditional independence. Thus, the partial correlation coefficient can be used for a nonparametric test of conditional independence.

Secondly, we present the approach of this paper, which has two key distinguishing features compared to the partial correlation approach: (i) the use of conditional ranks and (ii) complete removal of the marginal dependence. It is based on the new concept of the partial copula, which, like the partial correlation, is easily defined. Let  $X$  be a random variable on an arbitrary measurable space and let  $Y$  and  $Z$  be real-valued random variables with conditional distribution functions  $F_{Y|X}$  and  $F_{Z|X}$ . With

$$U = F_{Y|X}(Y|X) \quad \text{and} \quad V = F_{Z|X}(Z|X) \quad (2)$$

the partial copula pertaining to  $Y$  and  $Z$  given  $X$  is defined as the joint distribution function of  $(U, V)$  (cf. the ordinary copula pertaining to  $(Y, Z)$  is the joint distribution function of  $(Y, Z)$ ). We call transformation (2) the partial copula transform. If both  $Y$

and  $Z$  have continuous conditional distribution functions for every given value of  $X$ , then both  $U|X = x$  and  $V|X = x$  are uniformly distributed for all  $x$ , hence  $U \perp\!\!\!\perp X$  and  $V \perp\!\!\!\perp X$ . Therefore, by (1),  $Y \perp\!\!\!\perp Z|X$  implies  $U \perp\!\!\!\perp V$ . If we now apply the partial copula transform to iid data  $(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$ , i.e., setting

$$U_i = F_{Y|X}(Y_i|X_i) \quad \text{and} \quad V_i = F_{Z|X}(Z_i|X_i), \quad (3)$$

any test of ordinary independence for the  $(U_i, V_i)$  is a test for conditional independence for the  $(Y_i, Z_i)$  given the  $X_i$ .

In practice,  $F_{Y|X}$  and  $F_{Z|X}$  are typically unknown and need to be estimated. In Section 2 it is shown that for a wide variety of estimators, the asymptotic distribution of a broad class of test statistics in the form of a covariance between two functions on the marginals is unaffected by the estimation. Thus, a comprehensive range of tests of ordinary independence lead to an asymptotic test of conditional independence. The procedure is illustrated using a real data set and several tests of independence, including ones based on the covariance and Kendall's tau, and some tests with consistency against any alternative such as Hoeffding's test. Questions for further research are addressed in Section 4.

Note that, since  $U \perp\!\!\!\perp V$  does not imply  $Y \perp\!\!\!\perp Z|X$ , a test of the hypothesis  $U \perp\!\!\!\perp V$  cannot be consistent for all alternatives to the hypothesis  $Y \perp\!\!\!\perp Z|X$ : even if  $U \perp\!\!\!\perp V$  there may still be so-called three-variable interaction present. Fortunately, the testing for three-variable interaction seems in general a lot easier than the testing for conditional independence (Blum, Kiefer, & Rosenblatt, 1961), and it is reasonable to use a separate procedure for that. Using the partial copula transform, we can expect most power against alternatives with little three-variable interaction, in particular, alternatives for which the conditional copula, i.e., the joint distribution of  $(F_{Y|X}(Y|x), F_{Z|X}(Z|x))$ , does not vary much with  $x$ .

The following are some recent approaches with the same aim as the present paper. Su and White (2007, 2008) compare the conditional densities  $f_{Y|XZ}$  and  $f_{Y|X}$  using characteristic functions and Hellinger distances, while Bouezmarni et al. (2010) compare the conditional distribution functions  $F_{Y|XZ}$  and  $F_{Y|X}$ . Kernel methods are used to estimate the densities and distribution functions. The methods of Song (2009) and Huang (2010) are closer to the present paper's, in that they do not require estimation of the distribution of  $Y$  given both  $X$  and  $Z$ , but only given  $X$ . Huang uses Rényi type maximal correlations and differs from our approach in that it is not based on conditional ranks. Song's also uses the transformation (3), which he calls the Rosenblatt transform. Song and the present author found this independently, the present paper having appeared in different form as a technical report (Bergsma, 2004). Linton and Gozalo (1997) considered conditional independence tests based on Cramér von Mises and Kolmogorov-Smirnov criteria.

A brief selection of older approaches is as follows. Kendall (1942) introduced a partial version of Kendall's tau. However it is not clear how useful this coefficient can be in practice, as it is not necessarily zero under conditional independence unless certain restrictive conditions are met (Korn, 1984). Goodman (1959) and Gripenberg (1992) proposed a method based on another partial version of Kendall's tau, using the number of local

concordant and discordant pairs of observations. The easiest case for testing conditional independence is if  $X$  is categorical, with sufficiently many observations per category: for each category, a test of independence can be done, and these tests can be combined in various ways. If all three variables are categorical, log-linear techniques can be used (cf. Lauritzen, 1996, Agresti, 2002).

## 2 A practical approach to conditional independence testing

If  $F_{Y|X}$  and  $F_{Z|X}$  are known, the partial copula transform (3) can be applied and an arbitrary test of ordinary independence can be applied to the  $(U_i, V_i)$ . A permutation test would then be valid even for small samples. In Section 2.1 a general class of test statistics for independence testing is described. In practice, however,  $F_{Y|X}$  and  $F_{Z|X}$  are usually unknown and need to be estimated; in Section 2.2, we show that, under fairly unrestrictive conditions, this estimation does not affect the asymptotic behaviour of the test statistics given in Section 2.1. In Section 2.3, we illustrate the procedure on a real-data example using several tests of independence and provide some graphical displays.

### 2.1 A class of test statistics for ordinary independence

Many measures of association are of the form

$$\theta(Y, Z) = Es(Y_1, \dots, Y_r)t(Z_1, \dots, Z_r) \quad (4)$$

for some  $r \geq 1$  and where  $(Y_i, Z_i)$  are independent replications of  $(Y, Z)$ . Square integrability of  $s$  and  $t$  is a sufficient condition for the expectation to exist. If  $s$  and  $t$  have zero means, then  $Y \perp\!\!\!\perp Z$  implies  $\theta = 0$ . The sample ( $V$ -statistic) estimator which can be used as a test statistic for independence is

$$\hat{\theta}(Y, Z) = \frac{1}{n^r} \sum_{i_1=1}^n \dots \sum_{i_r=1}^n s(Y_{i_1}, \dots, Y_{i_r})t(Z_{i_1}, \dots, Z_{i_r}) \quad (5)$$

In the absence of distributional assumptions the best way to compute a  $p$ -value based on  $\hat{\theta}$  is usually the permutation test, but alternatively asymptotic theory is well-developed in certain cases (Hoeffding, 1948; Randles & Wolfe, 1979; Serfling, 1980).

The ordinary covariance and Kendall's tau are well-known association measures which can be written in the form (4). Spearman's rho can of course be written in that form as well, but we need not consider it because for rank data it is proportional to the covariance. Some coefficients of form (4) which are nonnegative and zero iff independence holds are the following. The first is Hoeffding's  $\Delta$ , which can be defined as

$$\Delta(Y, Z) = \frac{1}{4} E\phi((Y_1, Y_2, Y_3)\phi((Y_1, Y_4, Y_5)\phi((Z_1, Z_2, Z_3)\phi((Z_1, Z_4, Z_5)$$

where  $\phi(z_1, z_2, z_3) = I(z_1 \geq z_2) - I(z_1 \geq z_3)$ . Bergsma (2006) introduced

$$\kappa(Y, Z) = \frac{1}{4} E a(Y_1, Y_2, Y_3, Y_4) a(Z_1, Z_2, Z_3, Z_4)$$

and

$$\tau^*(Y, Z) = E \operatorname{sign}[a(Y_1, Y_2, Y_3, Y_4)] \operatorname{sign}[a(Z_1, Z_2, Z_3, Z_4)]$$

where

$$a(z_1, z_2, z_3, z_4) = |z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|$$

The latter is an extension of Kendall's tau introduced by Bergsma and Dassios (2010). Note that  $\Delta$ ,  $\kappa$  and  $\tau^*$  lead to tests consistent for all alternatives to independence.

## 2.2 Using estimated conditional distributions

In practice,  $F_{Y|X}$  and  $F_{Z|X}$  are typically unknown and need to be estimated. With estimates  $\hat{F}_{Y|X}$  and  $\hat{F}_{Z|X}$ , the partial copula transform yields

$$\hat{U}_i = \hat{F}_{Y|X}(Y_i|X_i) \quad \text{and} \quad \hat{V}_i = \hat{F}_{Z|X}(Z_i|X_i) \quad (6)$$

and substitution into (5) gives

$$\hat{\theta}(\hat{U}, \hat{V}) = \frac{1}{n^r} \sum_{i_1=1}^n \dots \sum_{i_r=1}^n s(\hat{U}_{i_1}, \dots, \hat{U}_{i_r}) t(\hat{V}_{i_1}, \dots, \hat{V}_{i_r})$$

Theorem 1 states that under the following easily satisfied conditions, the fact that we are using estimates  $(\hat{U}_i, \hat{V}_i)$  rather than true values  $(U_i, V_i)$  can be ignored.

- A1:  $s$  and  $t$  are continuous almost everywhere with respect to Lebesgue measure on  $[0, 1]^r$ ,
- A2:  $n^\alpha[\hat{\theta}(\hat{U}, \hat{V}) - \theta(U, V)] = O_p(1)$  for some positive  $\alpha$ ,
- A3: for almost all  $(x, y, z)$  and some positive  $\beta_1$  and  $\beta_2$ ,

$$\begin{aligned} n^{\beta_1}[\hat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)] &= O_p(1) \\ n^{\beta_2}[\hat{F}_{Z|X}(z|x) - F_{Z|X}(z|x)] &= O_p(1) \end{aligned}$$

with uniform convergence on all compact sets.

**Theorem 1** *Assume A1, A2, and A3 hold. Then*

$$n^\alpha[\hat{\theta}(\hat{U}, \hat{V}) - \theta(U, V)] = n^\alpha[\hat{\theta}(U, V) - \theta(U, V)] + O_p\left(n^{-\min(\beta_1, \beta_2)}\right)$$

**Proof:** Continuity of  $s$  and  $t$  (a.e.) implies continuity of  $s \times t$  (a.e.). Hence,

$$s(\hat{U}_{i_1}, \dots, \hat{U}_{i_r})t(\hat{V}_{i_1}, \dots, \hat{V}_{i_r}) - \theta(U, V) = \\ (s(U_{i_1}, \dots, U_{i_r})t(V_{i_1}, \dots, V_{i_r}) - \theta(U, V)) \left(1 + O_p(n^{-\min(\beta_1, \beta_2)})\right)$$

with probability 1. From this, the uniform convergence, and A2,

$$\begin{aligned} & n^\alpha[\hat{\theta}(\hat{U}, \hat{V}) - \theta(U, V)] \\ &= n^\alpha \left( \frac{1}{n^r} \sum_{i_1=1}^n \dots \sum_{i_r=1}^n s(\hat{U}_{i_1}, \dots, \hat{U}_{i_r})t(\hat{V}_{i_1}, \dots, \hat{V}_{i_r}) - \theta(U, V) \right) \\ &= n^\alpha \frac{1}{n^r} \sum_{i_1=1}^n \dots \sum_{i_r=1}^n [s(U_{i_1}, \dots, U_{i_r})t(V_{i_1}, \dots, V_{i_r}) - \theta(U, V)] \left[1 + O_p(n^{-\min(\beta_1, \beta_2)})\right] \\ &= n^\alpha[\hat{\theta}(U, V) - \theta(U, V)] \left[1 + O_p(n^{-\min(\beta_1, \beta_2)})\right] \\ &= n^\alpha[\hat{\theta}(U, V) - \theta(U, V)] + O_p(n^{-\min(\beta_1, \beta_2)}) \end{aligned}$$

with probability 1.  $\square$

Note that the proof is immediately adapted to  $U$ -statistic estimators. Assumptions A1 and A2 are satisfied for the measures of association mentioned in Section 2.1, i.e., the covariance, Kendall's tau, and  $\Delta$ ,  $\kappa$  and  $\tau^*$ . For Euclidean  $X$ , Nadaraya-Watson estimators are simple estimators of the marginal distribution functions satisfying A3 (under some additional restrictions). Alternatively, Hall, Wolff, and Yao (1999) introduced local linear and local logistic estimators, which improve the Nadaraya-Watson ones in certain respects. However, local linear estimators may not lead to a distribution function, and Hall et al. developed adjusted estimators which solve this problem.

### 2.3 Application with graphical illustration

We now apply the aforementioned tests to a real data example, illustrating the procedure with some graphical displays. Table 1 shows data on 35 consecutive patients under treatment for heart failure with the drug digoxin. The data are from Halkin, Sheiner, Peck, and Melmon (1975). Of medical interest is the hypothesis that digoxin clearance is independent of urine flow controlling for creatinine clearance, i.e.,  $Y \perp\!\!\!\perp Z|X$ . We computed (6) using the Nadaraya-Watson estimators

$$\hat{U}_i = \hat{F}_{Y|X}(y|x) = \frac{\sum_{i=1}^n K_1(|x - X_i|/h_1)I(Y_i \leq y)}{\sum_{i=1}^n K_1(|x - X_i|/h_1)} \quad (7)$$

and

$$\hat{V}_i = \hat{F}_{Z|X}(z|x) = \frac{\sum_{i=1}^n K_2(|x - X_i|/h_2)I(Z_i \leq z)}{\sum_{i=1}^n K_2(|x - X_i|/h_2)} \quad (8)$$

$X$	$Y$	$Z$	$X$	$Y$	$Z$	$X$	$Y$	$Z$
19.5	17.5	0.74	51.3	22.7	0.33	101.5	110.6	1.38
24.7	34.8	0.43	55.0	30.7	0.80	105.0	114.4	1.85
26.5	11.4	0.11	55.9	42.5	1.02	110.5	69.3	2.25
31.1	29.3	1.48	61.2	42.4	0.56	114.2	84.8	1.76
31.3	13.9	0.97	63.1	61.1	0.93	117.8	63.9	1.60
31.8	31.6	1.12	63.7	38.2	0.44	122.6	76.1	0.88
34.1	20.7	1.77	66.8	37.5	0.50	127.9	112.8	1.70
36.6	34.1	0.70	72.4	50.1	0.97	135.6	82.2	0.98
42.4	25.0	0.93	80.9	50.2	1.02	136.0	46.8	0.94
42.8	47.4	2.50	82.0	50.0	0.95	153.5	137.7	1.76
44.2	31.8	0.89	82.7	31.8	0.76	201.1	76.1	0.87
49.7	36.1	0.52	87.9	55.4	1.06			

Table 1: Digoxin clearance data. Clearances are given in  $\text{ml}/\text{min}/1.73\text{m}^2$ , urine flow in  $\text{ml}/\text{min}$ . Source: Halkin et al. (1975).

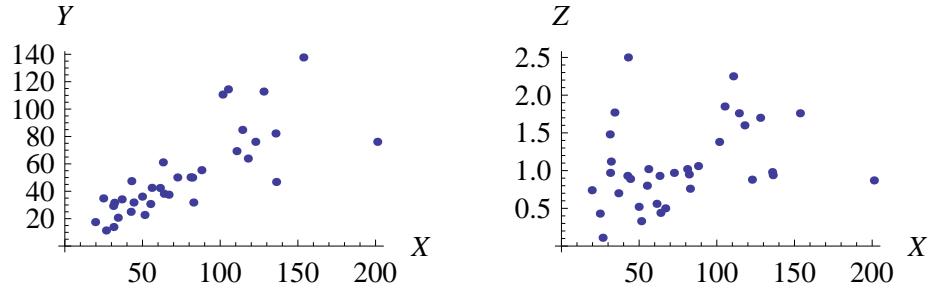
Note:  $X$ = Creatinine clearance,  $Y$ = digoxin clearance,  $Z$ = urine flow.

where  $h_t > 0$  is a bandwidth and  $K_t$  a kernel function ( $t = 1, 2$ ). We used a standard normal kernel and chose the bandwidth using Silverman's rule of thumb:  $h_1 = h_2 = 1.06\sigma_X n^{-1/5} = 22.48$ . See (Yu & Jones, 1998; Hall et al., 1999) and Koenker (2005) for other methods of estimation and bandwidth selection. In Figure 1(b) scatter plots are given of the pairs  $(X_i, \hat{U}_i)$  and  $(X_i, \hat{V}_i)$ . A visual inspection of both pictures confirms that the effect of  $X$  has been removed, that is, independence seems to hold. A scatterplot of the  $(\hat{U}_i, \hat{V}_i)$  is given in Figure 2 and some dependence is apparent, which can be tested for in various ways; in Table 2,  $p$ -values are given for tests based on several statistics. The  $p$ -values were approximated by the permutation test using  $10^5$  resamples. The simulations in Section 3 indicate that the tests are likely to be slightly liberal. Even taking that into account, there still appears to be good evidence for lack of conditional independence.

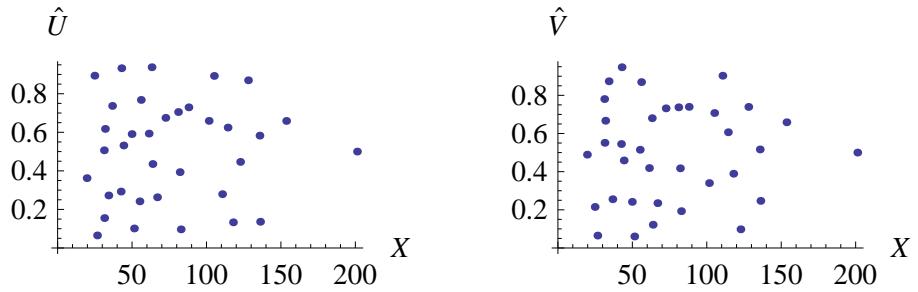
### 3 Simulation study

We try to answer the following two questions regarding conditional independence tests based on the partial copula:

1. How good are the type I error rates?
2. How much loss of power is there compared to the corresponding unconditional test of independence?



(a) Untransformed marginal distributions with association present



(b) Transformed marginal distributions with no apparent association

Figure 1: Illustration that the partial copula transformation of  $Y_i$ s into  $U_i$ s and  $Z_i$ s into  $V_i$ s removes marginal association. By (1), a test of independence for the  $(U_i, V_i)$ s is a test of conditional independence for the  $(Y_i, Z_i)$ s given the  $X_i$ s.

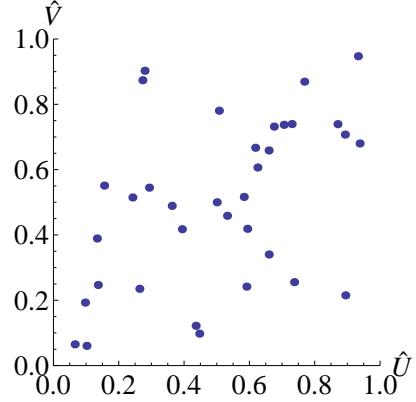


Figure 2: Partial copula transforms  $(\hat{U}_i, \hat{V}_i)$  for the digoxin data. Visual inspection suggests the presence of an association for the  $(\hat{U}_i, \hat{V}_i)$  which would imply a conditional association for the  $(Y_i, Z_i)$  given the  $X_i$ .

Test statistic	<i>p</i> -value
$\hat{\rho}(\hat{U}, \hat{V})$	.018
$\hat{\tau}(\hat{U}, \hat{V})$	.022
$\hat{\Delta}(\hat{U}, \hat{V})$	.107
$\hat{\kappa}(\hat{U}, \hat{V})$	.041
$\hat{\tau}^*(\hat{U}, \hat{V})$	.055

Table 2: Test statistics and associated *p*-values for the hypothesis  $Y \perp\!\!\!\perp Z | X$  for the digoxin data

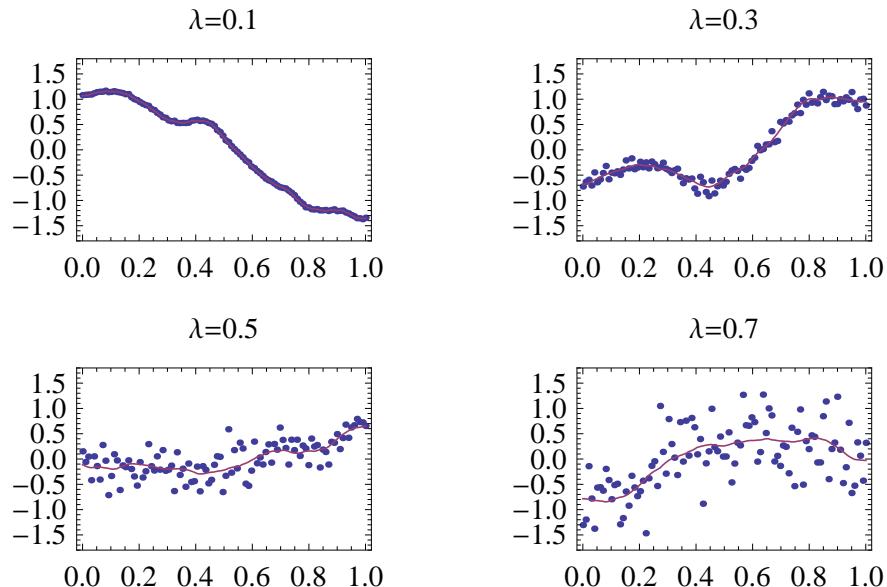


Figure 3: Randomly generated data with different noise to signal ratios  $\lambda$ . Here,  $\sigma_0 = 1$  and so  $\sigma_\varepsilon = \lambda/\sigma_0 = \lambda$ .

Data were generated according to the following model. We assumed

$$Y = g(x) + \varepsilon_Y(x) \quad \text{and} \quad Z = h(x) + \varepsilon_Z(x)$$

where the error pairs  $(\varepsilon_Y(x), \varepsilon_Z(x))$  follow a bivariate normal distribution with zero mean, (partial) correlation  $\rho_{YZ|X}$  and equal marginal standard deviations. The functions  $g$  and  $h$  were generated according to the integrated Wiener processes

$$g(x) = \sigma_0 \int_0^x W_{1t} dt \quad \text{and} \quad h(x) = \sigma_0 \int_0^x W_{2t} dt \quad (9)$$

where  $W_{1t}$  and  $W_{2t}$  are independent Wiener processes. Note that  $g(0) = h(0) = 0$ , which is arbitrary but does not affect the simulations. Further note that, with probability one,  $g$  and  $h$  are once differentiable. A ‘noise to signal ratio’ can be defined as  $\lambda = \sigma_\varepsilon/\sigma_0$ . In Figure 3, randomly generated (centered) curves and data are plotted with  $\sigma_0 = 1$  and  $\lambda \in \{0.1, .03, 0.5, 0.7\}$ . Finally, the values of  $x$  were chosen uniformly on  $[0, 1]$ .

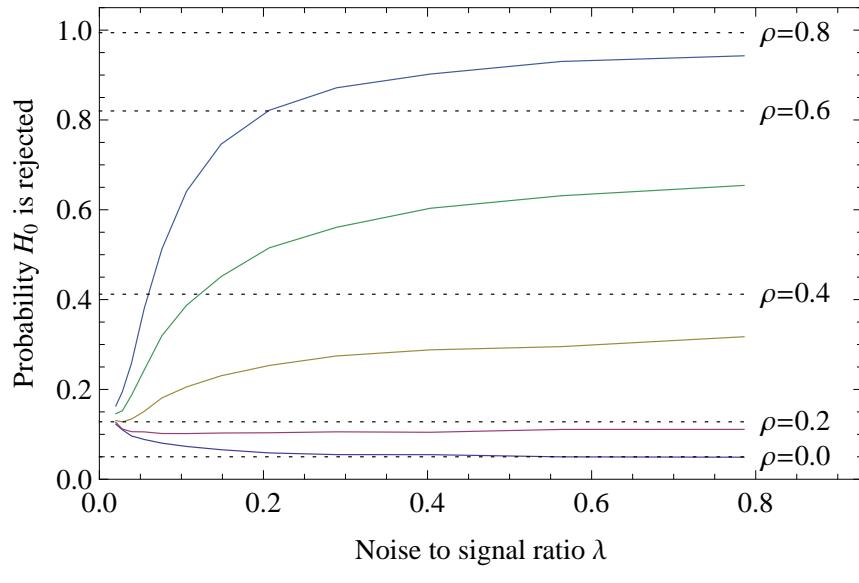
We simulated data according to the above model and estimated  $F_{Y|X}$  and  $F_{Z|X}$  using the Nadaraya-Watson estimators (7 and (8)). We took a standard normal kernel and chose bandwidths according to the formula  $h_1 = h_2 = 1.75\sqrt{\lambda/n}$ , which we empirically found to yield a kernel that gives a good approximation to the hat-matrix for the posterior mode for the above model. For each simulated data set, the  $p$ -value for the null hypothesis of conditional independence was obtained by performing a permutation test for independence on the  $\hat{U}_i$  and  $\hat{V}_i$ , where we used the sample correlation coefficient as a test statistic.

In Figure 4, it can be seen that for  $n = 20$  and  $n = 100$ , unless  $\lambda$  is very small, Type I errors error rates are close to nominal and the loss of power due to conditioning seems reasonable, especially for  $n = 100$ . For larger sample sizes the loss of power would become negligible. The method breaks down when  $\lambda$  becomes very small, which is to be expected as it leads to strong overfitting.

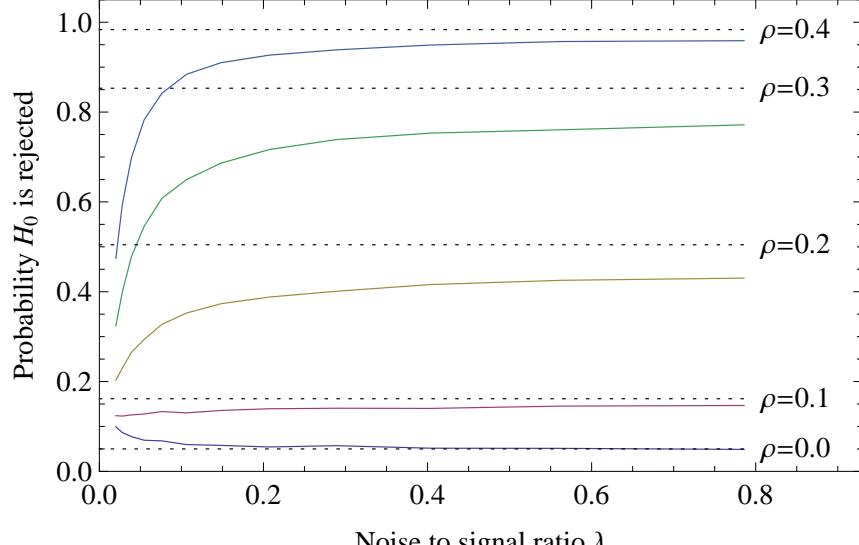
Figure 5 shows how the choice of bandwidth affects the Type I error rate. In particular for  $n = 100$ , a wide range of bandwidths give good error rates.

## 4 Remarks

The partial copula is a tool by means of which one’s favorite test of independence of appropriate form can be used for (rank) testing of conditional independence. Thus, if desired, consistency against arbitrary alternatives can be achieved by choosing an appropriate test statistic, as described in the paper. Furthermore, our method is much simpler than competing methods. For future research, the following questions might be investigated. Firstly, both the partial correlation and the partial copula approaches are based on removing association between response and control variables (see Section 1), but in seemingly quite different ways. This raises the question whether there is some more general approach to removing association. In particular, the case of multivariate response variables  $Y$  and  $Z$



(a)  $n = 20$



(b)  $n = 100$

Figure 4: Power curves and Type I error rates using  $\alpha = 0.05$  for several values of the partial correlation  $\rho$ . The dotted lines give the probabilities  $H_0$  is rejected for the corresponding test of marginal independence.

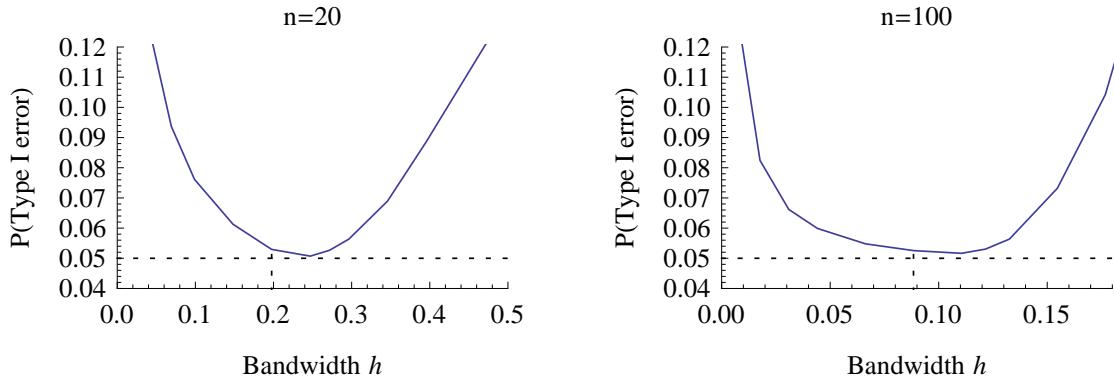


Figure 5: Type I error rates as a function of bandwidth for data generated using  $\lambda = 0.5$ . The vertical dotted lines represent the bandwidths used in Figure 4.

remains unsolved. Secondly, in its current form the partial copula approach does not work for discrete variables. We suspect a solution to this problem can be found by considering all possible orderings of tied observations, but this needs to be investigated further. Since excellent alternative methods are available when the explanatory variable is categorical Agresti (2002), this issue does not seem to be too pressing.

## References

Agresti, A. (2002). *Categorical Data Analysis, 2nd edition*. New York: Wiley.

Bergsma, W. P. (2004). Testing conditional independence for continuous random variables. *Eurandom technical report, 2004-048*, 1-19.

Bergsma, W. P. (2006). A new correlation coefficient, its orthogonal decomposition, and associated tests of independence. *arXiv:math/0604627v1 [math.ST]*.

Bergsma, W. P., & Dassios, A. (2010). An independence test for continuous and categorical ordinal data based on a sign correlation related to kendall's tau. *arXiv:1007.4259v1 [math.ST]*.

Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The annals of mathematical statistics*, 32, 485-498.

Bouezmarni, T., Rombouts, J. V., & Taamouti, A. (2010). A nonparametric copula based test for conditional independence with applications to granger causality. *Technical report*.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B*, 41, 1-31.

Goodman, L. A. (1959). Partial tests for partial tau. *Biometrika*, 46, 425-432.

Gripenberg, G. (1992). Partial rank correlations. *Journal of the American Statistical Association*, 87, 546-551.

Halkin, H., Sheiner, L. B., Peck, C. C., & Melmon, K. L. (1975). Determinants of the renal clearance of digoxin. *Clin. Pharmacol. Theor.*, 17, 385-394.

Hall, P., Wolff, R. C. L., & Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94(445), 154–163.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19, 293-325.

Huang, T. M. (2010). Testing conditional independence using maximal nonlinear conditional correlation. *Ann. Stat.*, 38, 2047-2091.

Kendall, M. G. (1942). Partial rank correlation. *Biometrika*, 32, 277-283.

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge: Cambridge University Press.

Korn, E. L. (1984). The ranges of limiting values of some partial correlations under conditional independence. *The American Statistician*, 38, 61-62.

Lauritzen, S. L. (1996). *Graphical models*. Oxford: Clarendon Press.

Linton, O., & Gozalo, P. (1997). Conditional independence restrictions: testing and estimation. *Cowles Foundation discussion paper 1140*.

Pearl, J. (2000). *Causality: modeling , reasoning, and inference*. New York: Cambridge University Press.

Randles, R. H., & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*. New York: Wiley.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley-Blackwell.

Song, K. (2009). Testing conditional independence via Rosenblatt transforms. *Ann. Stat.*, 37, 4011-4045.

Su, L., & White, H. (2007). A consistent characteristic-function-based test for conditional independence. *J. of Econometrics*, 141, 807-834.

Su, L., & White, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory*, 24, 829-864.

Wermuth, N., & Cox, D. (1996). *Multivariate dependencies: Models, analysis and interpretation*. London: Chapman and Hall.

Yu, K., & Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.*, 93(441), 228–237.